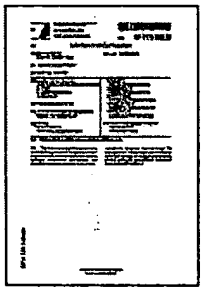
EP0786519A2: Staphylococcus aureus polynucleotides and sequences

EP European Patent Office (EPO)

A2 Publ. of Application without search report



High Resolution

Low Resolution

3271 pages

Kunsch, Charles A.
Choi, Gil H.
Barash, Steven C.
Dillon, Patrick J.
Fannon, Michael R.
Rosen, Craig A.

HUMAN GENOME SCIENCES, INC.
News, Profiles, Stocks and More about this company

July 30, 1997 / Jan. 7, 1997

EP1997000100117

C12N 15/00;

C07K14/31; C12N15/10C; G06F17/30A2;

Jan. 5, 1996 **US1996000009861**

Show legal status actions

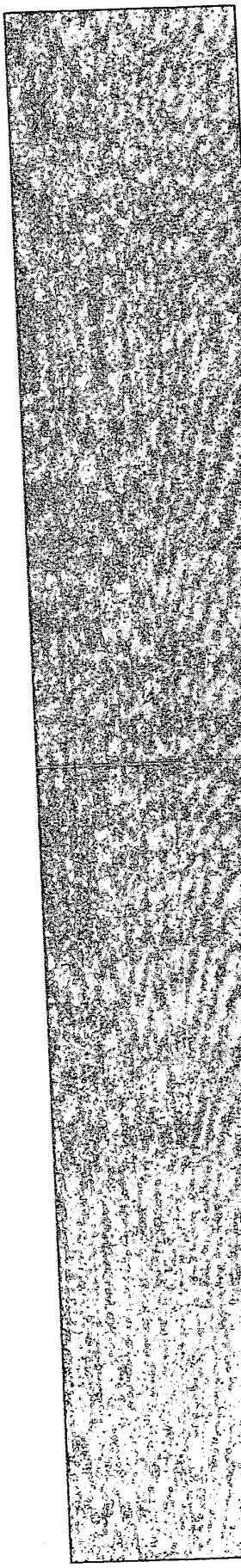
AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LI, LU, MC, NL, PT, SE

The present invention provides polynucleotide sequences of the genome of Staphylococcus aureus, polypeptide sequences encoded by the polynucleotide sequences, corresponding polynucleotides and polypeptides, vectors and hosts comprising the polynucleotides, and assays and other uses thereof. The present invention further provides polynucleotide and polypeptide sequence information stored on computer readable media, and computer-based systems and methods which facilitate its use.

Show known family members

Expand full description

+ **Human Health and *S. Aureus***
+ ***Burns***



+ ***Cellulitis***
+ ***Eyelid infections***
+ ***Food poisoning***
+ ***Joint infections***
+ ***Osteomyelitis***
+ ***Skin infections***
+ ***Surgical Wound Infections***
+ ***Scalded Skin Syndrome***
+ ***Toxic Shock Syndrome***
+ ***Nocosomial Infections***
+ ***Resistance to drugs of S. aureus strains***
+ **Molecular Genetics *of Staphylococcus Aureus***
- **COMPUTER RELATED EMBODIMENTS**

The nucleotide sequences provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide sequence of SEQ ID NOS:1-5,191 may be "provided" in a variety of mediums to facilitate use thereof. As used herein, Òprovided" refers to a manufacture, other than an isolated nucleic acid molecule, which contains a nucleotide sequence of the present invention; *i.e.*, a nucleotide sequence provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide of SEQ ID NOS:1-5,191. Such a manufacture provides a large portion of the *Staphylococcus aureus* genome and parts thereof (*e.g.*, a *Staphylococcus aureus* open reading frame (ORF)) in a form which allows a skilled artisan to examine the manufacture using means not directly applicable to examining the *Staphylococcus aureus* genome or a subset thereof as it exists in nature or in purified form.

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD- ROM; electrical storage media such as RAM and ROM; and hybrids of these categories, such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. Likewise, it will be clear to those of skill how additional computer readable media that may be developed also can be used to create analogous manufactures having recorded thereon a nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently know methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure

will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially- available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data-processor structuring formats (*e.g.*, text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Thus, by providing in computer readable form the nucleotide sequences of SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a sequence of SEQ ID NOS:1-5,191 the present invention enables the skilled artisan routinely to access the provided sequence information for a wide variety of purposes.

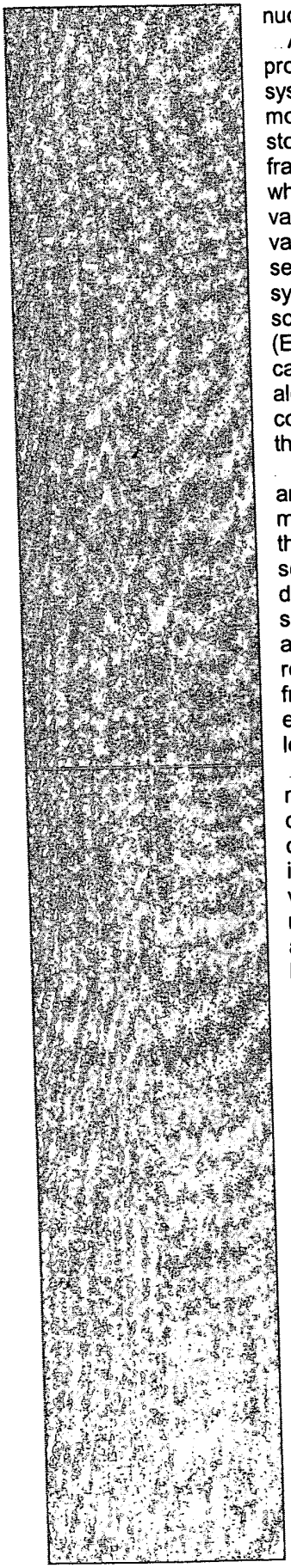
The examples which follow demonstrate how software which implements the BLAST (Altschul *et al.*, J. Mol. Biol. 215:403410 (1990)) and BLAZE (Brutlag *et al.*, Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system was used to identify open reading frames (ORFs) within the *Staphylococcus aureus* genome which contain homology to ORFs or proteins from both *Staphylococcus aureus* and from other organisms. Among the ORFs discussed herein are protein encoding fragments of the *Staphylococcus aureus* genome useful in producing commercially important proteins, such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

The present invention further provides systems, particularly computer-based systems, which contain the sequence information described herein. Such systems are designed to identify, among other things, commercially important fragments of the *Staphylococcus aureus* genome.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention.

As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means.

As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the



nucleotide sequence information of the present invention.

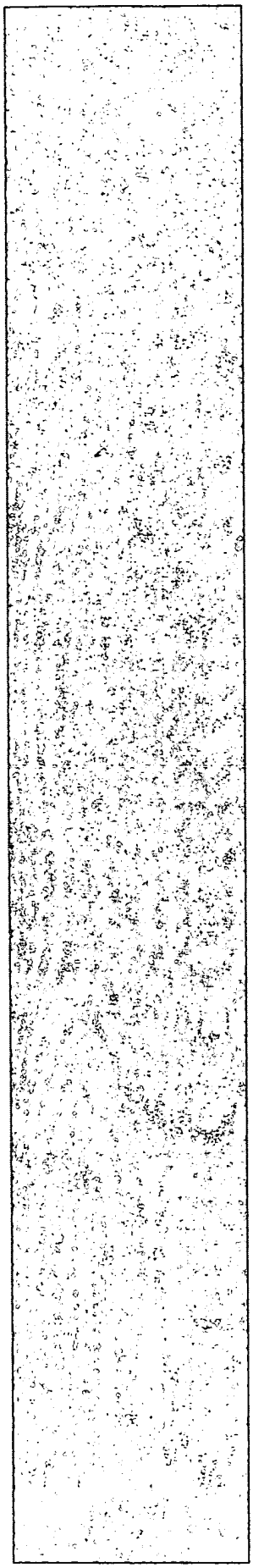
As used herein, "search means" refers to one or more programs which are implemented on the computer- based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the present genomic sequences which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymic active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the *Staphylococcus aureus* genomic sequences possessing varying degrees of homology to the target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments of the *Staphylococcus aureus* genome. In the present examples, implementing software which implement the BLAST and BLAZE algorithms, described in Altschul *et al., J. Mol. Biol.* 215: 403-410 (1990), was used to identify open reading frames within the *Staphylococcus aureus* genome. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used



as the search means for the computer-based systems of the present invention. Of course, suitable proprietary systems that may be known to those of skill also may be employed in this regard.

Figure 1 provides a block diagram of a computer system illustrative of embodiments of this aspect of present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented as random access memory, RAM) and a variety of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, *etc.* A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, *etc.*) containing control logic and/or data recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the data from the removable medium storage device 114, once it is inserted into the removable medium storage device 114.

A nucleotide sequence of the present invention may be stored in a well known manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. During execution, software for accessing and processing the genomic sequence (such as search tools, comparing tools, *etc.*) reside in main memory 108, in accordance with the requirements and operating parameters of the operating system, the hardware system and the software program or programs.

± BIOCHEMICAL EMBODIMENTS
± ILLUSTRATIVE USES OF COMPOSITIONS OF THE INVENTION
± 1. Biosynthetic Enzymes
± 2. Generation of Antibodies
± 3. Diagnostic Assays and Kits
± 4. Screening Assay for Binding Agents
± 5. Pharmaceutical Compositions and Vaccines
± 6. Shot-Gun Approach to Megabase DNA Sequencing
± ILLUSTRATIVE EXAMPLES
± LIBRARIES AND SEQUENCING
± 1. Shotgun Sequencing Probability Analysis
± 2. Random Library Construction
± 3. Random DNA Sequencing
± 4. Protocol for Automated Cycle Sequencing
± INFORMATICS
± 1. Data Management
± 2. Assembly
± 3. Identifying Genes
± ILLUSTRATIVE APPLICATIONS
± 1. Production of an Antibody to a *Staphylococcus aureus* Protein
± 2. Monoclonal Antibody Production by Hybridoma Fusion
± 3. Polyclonal Antibody Production by Immunization
± 3. Preparation of PCR Primers and Amplification of DNA
± 4. Gene expression from DNA Sequences

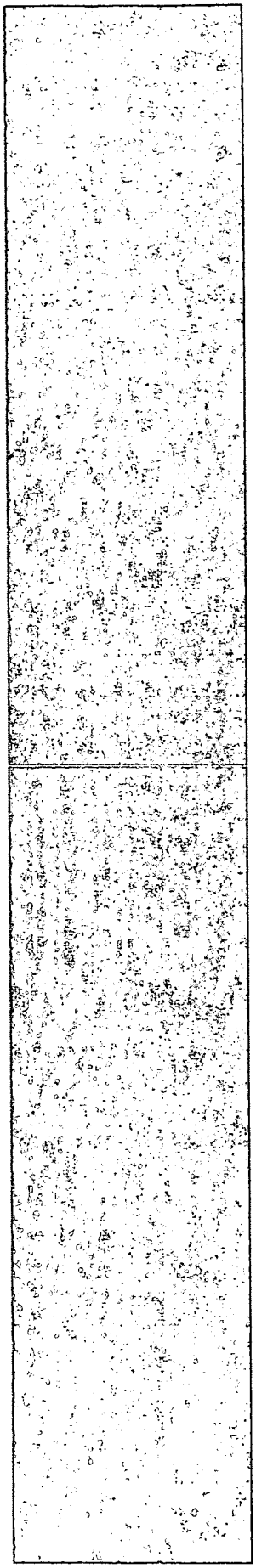
Claims
[Hide claims]

1. Computer readable medium having recorded thereon a nucleotide sequence of the Staphylococcus aureus genome as depicted in SEQ ID NOS:1-5,191, a representative fragment thereof or a nucleotide sequence at least 95 % identical to a nucleotide sequence depicted in SEQ ID NOS:1-5,191.

2. Computer readable medium having recorded thereon any one of the fragments of SEQ ID NOS:1-5,191 depicted in Tables 2 and 3 or a degenerate variant thereof.

3. The computer readable medium of claim 1, wherein said medium is selected from the group consisting of a floppy disc, a hard disc, random access memory (RAM), read only memory (ROM), and CD-ROM.

4. The computer readable medium of claim 3, wherein said medium is selected from the group consisting of a floppy disc, a hard disc, random access memory (RAM),



read only memory (ROM), and CD-ROM.

5. A computer-based system for identifying fragments of the *Staphylococcus aureus* genome of commercial importance comprising the following elements:

- (a) a data storage means comprising the nucleotide sequence of SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95% identical to a nucleotide sequence of SEQ ID NOS:1-5,191;
- (b) search means for comparing a target sequence to the nucleotide sequence of the data storage means of step (a) to identify homologous sequence(s), and
- (c) retrieval means for obtaining said homologous sequence(s) of step (b).

6. A method for identifying commercially important nucleic acid fragments of the *Staphylococcus aureus* genome comprising the step of comparing a database comprising the nucleotide sequences depicted in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95% identical to a nucleotide sequence of SEQ ID NOS:1-5,191 with a target sequence to obtain a nucleic acid molecule comprised of a complementary nucleotide sequence to said target sequence, wherein said target sequence is not randomly selected.

7. A method for identifying an expression modulating fragment of *Staphylococcus aureus* genome comprising the step of comparing a database comprising the nucleotide sequences depicted in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95% identical to the nucleotide sequence of SEQ ID NOS:1-5,191 with a target sequence to obtain a nucleic acid molecule comprised of a complementary nucleotide sequence to said target sequence, wherein said target sequence comprises sequences known to regulate gene expression.

8. A protein-encoding nucleic acid fragment of the Staphylococcus aureus genome,

wherein said fragment comprises the nucleotide sequence of any one of the fragments of SEQ ID NOS:1-5,191 depicted in Tables 2 and 3 or a nucleotide sequence at least 95% identical to such a nucleotide sequence, or a degenerate variant of any of the aforementioned sequences.

9. The nucleic acid fragment of claim 8 which is DNA.

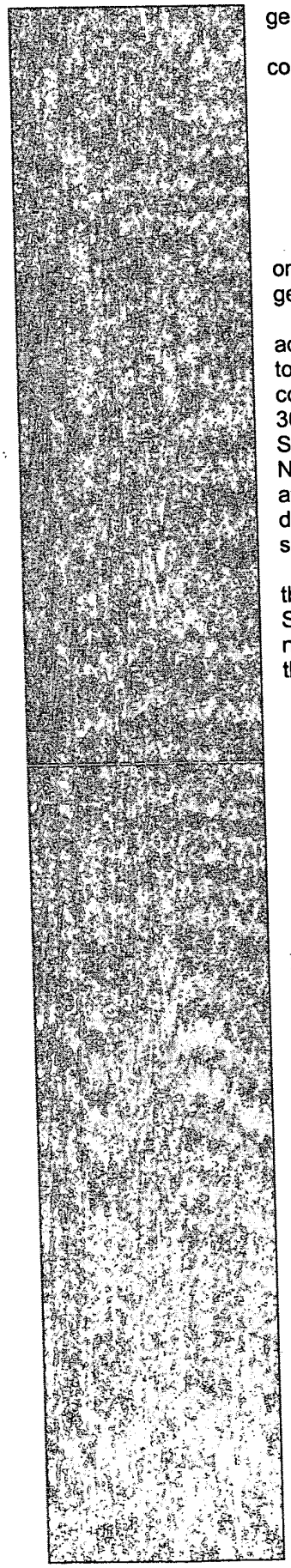
10. The nucleic acid fragment of claim 8 which is RNA.

11. A vector comprising a fragment of claim 8.

12. A fragment of the Staphylococcus aureus genome, wherein said fragment modulates the expression of an operably liked open reading frame, wherein said fragment consists of the nucleotide sequence from about 10 to 200 bases in length which is 5' to any one of the open reading frames depicted in Tables 2 and 3 or a nucleotide sequence at least 95% identical to such a nucleotide sequence or a degenerate variant of any of the aforementioned sequences.

13. A vector comprising a fragment of claim 12.

14. A organism which has been altered to contain any one of the fragments of the Staphylococcus aureus



genome of claim 8.

15. A method for producing a polypeptide in a host cell comprising the steps of:

- (a) incubating an organism of claim 14 under conditions where said fragment is expressed to produce said protein, and
- (b) isolating said protein.

16. An organism which has been altered to contain any one of the fragments of the Staphylococcus aureus genome of claim 12.

17. A method for regulating the expression of a nucleic acid molecule comprising the step of covalently attaching to said nucleic acid molecule a nucleic acid molecule consisting of the nucleotide sequence from about 30 to 300 bases 5' to any one of the fragments of the Staphylococcus aureus genome depicted in Seq ID Nos:1-5,191 and Tables 2 and 3 or a nucleotide sequence at least 95% identical to such a nucleotide sequence or a degenerate variant of any of the aforementioned sequences.

18. A nucleic acid molecule being a homolog of any of the fragments of the Staphylococcus aureus genome of SEQ ID NOS:1-5,191 and Tables 2 and 3, wherein said nucleic acid molecule is produced by a process comprising the steps of:

- (a) screening a genomic DNA library using as a probe a target sequence defined by any of SEQ ID NOS:1-5,191 and Tables 2 and 3, including fragments thereof;
- (b) identifying members of said library which contain sequences that hybridize to said target sequence;
- (c) isolating the nucleic acid molecules from said members identified in step (b).

19. A DNA molecule being a homolog of any one of the fragments of the Staphylococcus aureus genome of SEQ ID NOS:1-5,191 and Tables 2 and 3, wherein said nucleic acid molecule is produced by a process comprising the steps of:

- (a) isolating mRNA, DNA, or cDNA produced from an organism;
- (b) amplifying nucleic acid molecules whose nucleotide sequence is homologous to amplification primers derived from said fragment of said *Staphylococcus aureus* genome to prime said amplification;
- (c) isolating said amplified sequences produced in step (b).

20. A polypeptide encoded by a fragment of claim 8.

21. An antibody which selectively binds to any one of the polypeptides of claim 20.

22. A kit for analyzing samples for the presence of polynucleotides derived from Staphylococcus aureus, comprising
at least one polynucleotide containing a nucleotide sequence of any one of the fragments SEQ ID NOS:1-5,191 depicted in Tables 2 and 3 or a nucleotide

sequence at least 95% identical thereto or a degenerate variant of any of the aforementioned sequences, that will hybridize to a staphylococcus aureus polynucleotide under stringent hybridization conditions, and a suitable container.

23. A Staphylococcus aureus polypeptide comprising an amino acid sequence identical to an amino acid sequence selected from the group consisting of SEQ ID NOS:5,192 to 5,255 or comprising an amino acid sequence having at least 95% identity to such a sequence.

24. A Staphylococcus aureus polypeptide antigen comprising at least one epitope derived from a Staphylococcus aureus polypeptide selected from the group consisting of SEQ ID NOS:5,192 to 5,255.

25. A polypeptide comprising at least one epitope encoded by a Staphylococcus aureus amino acid sequence selected from the group consisting of the epitopic sequences listed in Table 4.

26. The polypeptide of claim 24 or 26, wherein said polypeptide is fixed to a solid phase.

27. A diagnostic kit for detecting Staphylococcus aureus infection comprising

- (a) an isolated polypeptide antigen of claim 24, and
- (b) means for detecting the binding of an antibody contained in a biological fluid to said antigen.

28. A vaccine composition comprising a polypeptide of claim 24 present in a pharmaceutically acceptable carrier.

29. A method of vaccinating an individual against Staphylococcus aureus infection comprising, administering to an individual the vaccine composition of claim 28.

Other Abstract Info: CHEMABS 127(16)215968Y CHEMABS 127(16)215968Y DERABS C1997-374922 DERABS C1997-374922

Foreign References:

Show the 2 patents that reference this one